Surgical Resident Accuracy in Predicting Their ABSITE Score

LaShondria Simpson-Camp, MD, Edward A. Meister, PhD, Stephen Kavic, MD

ABSTRACT

Background: The American Board of Surgery In-Training Examination (ABSITE) is given to all surgical residents as an assessment tool for residents and their programs in preparation for the American Board of Surgery qualifying and certifying examinations. Our objective was to ascertain how well surgical residents could predict their percentile score on the ABSITE using two predictor measures before and one immediately after the examination was completed.

Methods: A survey was given to surgical residents in postgraduate year(s) (PGY) 2 through 5 as well as to research residents in November and December 2011, and immediately after the examination in January 2012, to ascertain their predicted ABSITE scores. Thirty-one general surgery residents were measured consisting of PGY-2 (22%), PGY-3 (19.4%), PGY-4 (19.4%), and PGY-5 (12.9%), and research residents 25.8%.

Results: Mean prediction scores were consistently higher than actual examination scores for both junior and senior examination takers, with senior examination predictions exhibiting the highest proportion of variation on the actual examination score. Stratified linear regression analysis showed little predictive significance of all 3 examination predictions and actual score, except for the senior examination predictions in November 2011 (t test $= 2.521$, $P = .027$). We found no statistically significant difference in the proportion of residents overestimating or underestimating their predicted score. Secondary analysis using a linear regression model shows that 2011 scores were a statistically significant predictor of 2012 scores (overall F $= 13.258$, $P = .001$, $R^2 = 0.31$) for both junior and senior examinations.

Conclusion: General surgery residents were not able to accurately predict their ABSITE score; however, the previous year's actual scores were found to have the most predictive value of the next year's actual scores.

Key Words: ABSITE, In-training examination, Surgical education.

Department of Surgery, Hendrick Medical Center, Abilene, Texas, USA (Dr. Simpson-Camp).

Department of Medicine, University of Arizona, Tucson, Arizona, USA (Dr. Meister).

Department of Surgery, University of Maryland School of Medicine, Baltimore, Maryland, USA (Dr. Kavic).

Address correspondence to: Stephen Kavic, MD, Department of Surgery, University of Maryland School of Medicine, 22 S. Greene St., S4B09, Baltimore, MD 21201. Telephone: (410) 328-7592, Fax: (410) 328-5919, E-mail: skavic@smail.umaryland.edu

INTRODUCTION

Within the last ten years, there has been considerable interest in the predictive value of in-service training examinations of medical residents and their outcomes on national board certification examinations.[1] The American Board of Surgery In-Training Examination for resident surgeons, the ABSITE, is no exception. Certainly, to some extent the ABSITE is a monitoring tool to see how well surgical training residents and programs are at preparing the American Board of Surgery's residents to take the Qualifying Examination and the General Surgery Certifying Examination. Current research indicates a strong predictive component of the ABSITE scores and passing or failing the certifying board examination.

Preparing for the ABSITE involves considerable time and effort, and surgery residents are in a continuous mode of preparation and self-appraisal before taking the in-service examinations. The issue is critical for both residents and program directors. If a resident's self-appraisal is inaccurate, especially in overestimating their results, then more needs to be done to facilitate better examination preparation. Previously, family medicine residents were studied to determine how well they were able to predict their in-training examinations. Parker and colleagues[2] conducted a survey of 17 family residency programs in Texas and Oklahoma, using a 100-point visual analog scale to measure each resident's predictive score. They found that residents' predicted scores correlated poorly with their actual performance scores. Men had the highest correla-

tion of 0.293 ($P = .09$) for the content area of obstetrics, whereas other correlations were much lower. Recently, Jones and colleagues undertook a similar study to determine the predictive accuracy of internal medicine residents to assess their medical knowledge in anticipation of taking the in-training examination.[3] The Jones study included 26 internal medicine residents who were asked in a written survey to predict their percentile score 1 week before and 1 week after the in-training examinations. Their results indicated that the residents were poor predictors of their actual percentile scores: only 31% of residents were within 10 points of their predictions, and taking the examination did not improve predictive ability.

Our objective was to ascertain how well surgery residents could predict their percentile score on the ABSITE using two predictor measures before and one immediately after the examination was completed.

METHODS

Surgery residents in training and surgery research residents were prospectively surveyed to obtain their ABSITE percentile prediction score for their January 2012 in-service examination. The survey was given in November and December 2011, and in January 2012 immediately after the examination. The principal investigator was blinded to the predictions. There are two ABSITE examinations administered. The junior examination is given to residents during postgraduate year(s) (PGY) 1 and 2, as well as to research residents, and consists of 60% basic science questions and 40% clinical management questions. The senior examination, given to residents during PGYs 3–5, consists of 80% clinical management questions. Our study included surgery residents PGYs 2–5 and research residents. In our statistical analysis, we calculated 3 precision scores (PS) for each resident, taking the predicted percentile scores and subtracting the actual scores for the 3 prediction periods—2 months prior, 1 month prior, and immediately after the actual examination—as follows:

PS = Predicted Percentile Score − Actual Percentile Score

A positive PS indicates the resident overestimated their performance, and a negative PS indicates they underestimated their performance. Freidman's analysis of variance for repeated measures was used to assess the change in PS scores over time, stratified by the ABSITE examination for juniors and seniors. Multiple linear regression modeling was undertaken with the actual score as the dependent variable, and November 2011, December 2011, and January 2012 posttest predictions stratified by examination

Table 1.
Mean Scores and Prediction Scores by Exam Types and Month

Test Groups	Actual Score	Nov 2011	Dec 2011	Jan 2012
Junior test				
Mean	62.67	71.60	66.13	65.93
CV	35.08%	17.22%	24.01%	22.49%
Senior test				
Mean	37.50	73.50	65.94	64.50
CV	70.71%	20.09%	43.86%	37.84%

CV = coefficient of variation.

groups. Pearson's product-moment correlation analysis was used to assess bivariate associations, and $P \le .05$ was considered statistically significant. We calculated the proportion of residents whose prediction was within 1 SD of the actual score (based on their PS z-score) to adjust for the whole number bias because they were likely not to pick numbers between 1 and 4 or between 6 and 9, and were likely to round up or down to the nearest tenth. Data were collected, compiled in a spreadsheet, and analyzed using SPSS version 19.0 (IBM, Armonk, NY).

RESULTS

The scores of 31 general surgery residents were measured in this study. The composition by years was as follows: PGY-2 (22%), PGY-3 (19.4%), PGY-4 (19.4%), PGY-5 (12.9%), and research residents (25.8%). The examination proportions were evenly divided: 48.4% (n = 15) took the junior examination, and 52.6% (n = 16) took the senior examination. Mean predicted scores were consistently higher than actual scores for both the junior and senior examinations, and the senior examination predictions had a much higher coefficient of variability than was evident in the junior examination predictions, particularly for their actual scores (**Table 1**).

Overall bivariate correlations of predictions with actual score were mean correlations of $r = 0.489$ for the junior examination and $r = 0.626$ ($P < .05$) for the senior examination. For residents taking the junior examinations, we observed an increasing trend in improving correlations with actual scores, whereas for the senior examinations, the trend noted was a decreasing correlation between predictions and actual scores (**Figure 1**). For all 3 senior examination predictions, there was a statistically significant correlation of percentile predictions with actual scores ($r = 0.693$, $r = 0.632$, and $r = 0.560$; $P < .001$),
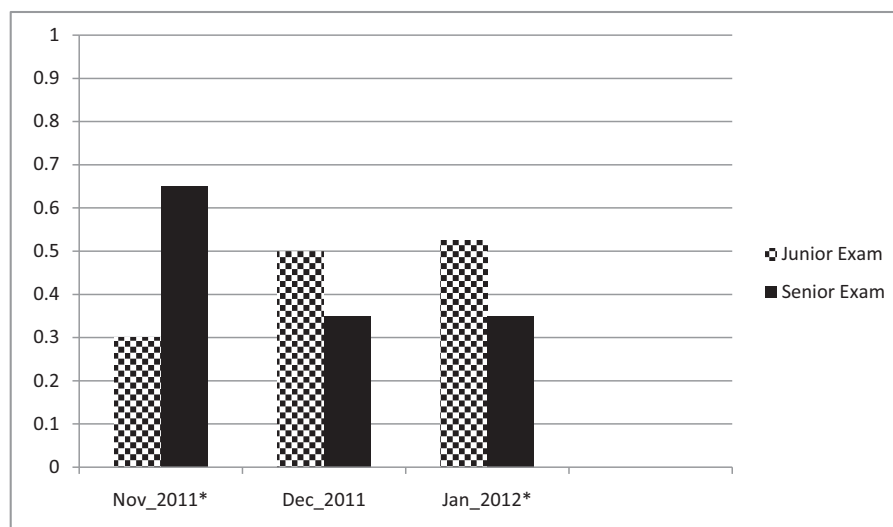
Figure 1. Correlation of percentile predictions with actual January 2012 scores by examination level.
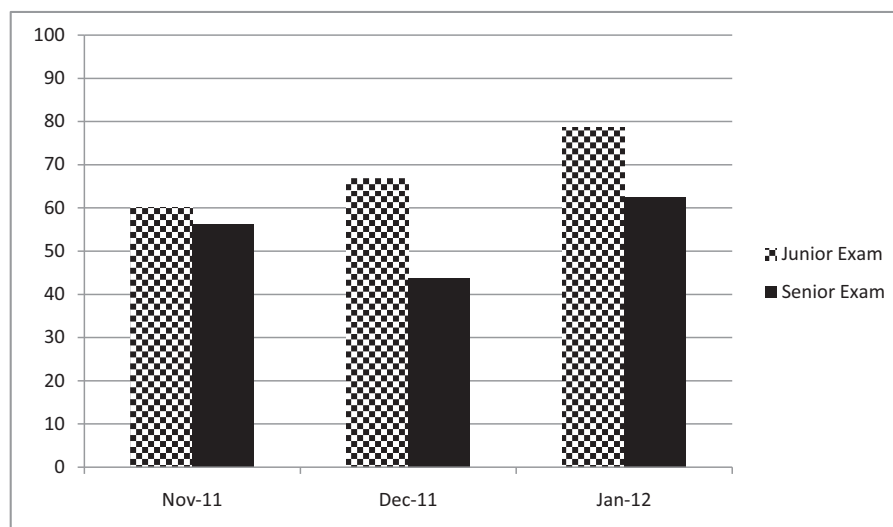

Figure 2. Percentage of predictions within 1 SD of actual score.

whereas the junior examination correlations were only statistically significant with the predictions immediately before and after the actual examinations ($r = 0.523$ and $r = 0.636$, respectively; $P < .001$).

However, when examining the proportion of individual z-scores that were within 1 SD of the actual score, residents taking the junior examination tended to have projected scores closer to their actual scores than those taking the senior examination. A mean of 68.43% of those taking the junior examination had predictions within 1 SD of their predicted score, whereas a mean of 52.83% of those taking the senior examination had prediction scores

within 1 SD of their predicted score; yet this difference was not statistically significant ($P > .05$).

Figure 2 shows that residents taking the junior examination had a higher proportion of predictions that were within 1 SD of their actual scores, and this trend continued through all predictions. Residents taking the senior examination were inconsistent in their examination prediction scores, yet after taking the examination, their accuracy improved significantly.

Overall multiple linear regression stratified by groups indicated that residents were poor predictors of their actual

percentile scores. **Table 2** shows that all of the pretest prediction coefficients from November and December 2011, and the posttest prediction coefficients of January 2012 were not statistically significant for predicting the January 2012 actual score, with the exception of The November 2011 predictions those of the residents taking the senior examination ($P = .027$).

Secondary analysis using a linear regression model to determine the predictive merit of 2011 for the 2012 scores shows that 2011 scores were a statistically significant predictor of 2012 scores for both groups (junior examination: F = 7.458, $P = .017$, $R^2 = 0.365$; senior examination: F = 7.642, $P = 0.015$, $R^2 = 0.353$). Finally, χ^2 analysis found no statistically significant difference in the proportion of residents who underestimated or overestimated their examination scores based on the two predictions before and after the actual examination.

DISCUSSION

The capacity of the surgery residents to predict their ABSITE score was poor as is evident by the low overall linear correlation coefficients for both the junior and senior examination versions, as is the lack of any significant beta coefficients in the overall linear regression model.

Our empirical analysis indicated that residents taking the junior examination increased their predictive accuracy, feeling perhaps more confident in their ability after taking the examination. This trend was the opposite for those taking the senior examination, possibly underlying a substantial drop in confidence in their abilities immediately after the examination. It appears that senior residents lost confidence the more proximate they were to the actual examination date and underestimated their performance, whereas the junior residents gained confidence the more proximate to the

actual examination by increasing their predicted scores. Our results among internal medicine residents are similar to the research by Jones and colleagues; however, our residents who took the junior examination had a greater predictive accuracy after taking the examination. The surgery residents in our study also had much higher correlations between actual and predicted scores compared with those found among family medicine residents in the study by Parker and colleagues. On secondary analysis, the strongest predictor of the current year's ABSITE score was the previous year's score. Both juniors and seniors did not differ in their level of over- or underestimation of their scores on all pre- and posttests, indicating similar levels of prediction confidence.

Our study results are limited to a single institution of surgical residency training, thus external generalization of the results are somewhat limited. However, our study is the first to specifically focus on surgery residents taking the ABSITE examination and to compare predictions with actual results longitudinally, which investigates the effects of distal versus proximal score projections and the actual scores.

Self-assessment by surgery residents on predictions of their ABSITE scores is an important indicator of their ability to be lifelong learners and demonstrate their confidence in their medical knowledge. In-training evaluations are benchmarks to assist residents in preparation for the qualifying and board examinations. In-training examinations are not trivial because the performance on the ABSITE is predictive of successfully passing the ABS board examination.[4] The goal of all evaluations should be to encourage learning in a positive manner. Recent research indicates that the usefulness of medical training examinations to enhance learning is dependent on the examination taker's perception

Table 2.
Stratified Linear Regression Analysis of Actual Score-Dependent Variable

Groups	Tests	Unstandardized Coefficients	Standard Error	Standardized Coefficients	t Test	P Value
Junior	Nov 2011	0.020	0.577	0.011	0.035	.973
	Dec 2011	0.086	1.058	0.062	0.081	.937
	Jan 2012	0.681	1.095	0.459	0.621	.547
Senior	Nov 2011	1.256	0.498	0.7	2.521	.027
	Dec 2011	−0.205	0.343	−0.224	−0.599	.561
	Jan 2012	0.205	0.36	0.188	0.569	.58

of the process; thus, all programs of resident training need to be open to considering changes to improve the examination experience.[5]

CONCLUSION

Surgery residents taking ABSITE examinations were poor in estimating their actual percentile scores. However, the previous year's scores showed evidence of predicting the following year's actual score regardless of examination level.

References:

1. Ozuah PO. Predicting resident's performance: A prospective study. *BMC Med Educ*. 2002;2:7.

2. Parker RW, Alford C, Passmore C. Can family medical residents predict their performance on the in-training examination? *Fam Med*. 2004;36(10):705–709.

3. Jones R, Panda M, Desbiens N. Internal medicine residents do not accurately assess their medical knowledge. *Adv Health Sci Educ*. 2008;13:463–468.

4. Virgilio C, Yaghoubian A, Kaji A, et al. Predicting performance on the American Board of Surgery qualifying and certifying examinations. *Arch Surg*. 2010;145(9):852–856.

5. Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participant's perceptions of the process. *Adv Health Sci Educ*. 2012;17:183–194.